

ANÁLISIS DE BIG DATA

Código: BIG-006

Propuesta de Valor: BIG DATA

Duración: 45 Horas Académicas



El objetivo de este curso es ofrecer una capacitación intermedia sobre los principales métodos, tecnologías y herramientas utilizados actualmente para analizar conjuntos complejos de datos, denominados Big Data. Se define como big data aquellos conjuntos de datos cuyo procesamiento no se puede abordar mediante técnicas de computación o análisis convencionales, precisándose métodos más avanzados. Para ello, se describirán los fundamentos teóricos para el tratamiento y procesamiento de big data, incluyendo aspectos tanto de infraestructura computacional (ingeniería de datos), como de administración de datos (bases de datos no convencionales) así como métodos avanzados de análisis. Nuestro programa está orientado a la capacitación de personas interesadas en trabajar en análisis de datos, y en particular de análisis que involucran la evaluación de volúmenes muy grandes de datos en empresas y organizaciones. Se hará especial hincapié en el uso de aplicaciones basadas en lenguajes analíticos como R y Python. Los cursos también introducirán paradigmas de almacenamiento y procesamiento de datos como MapReduce o noSQL, y sus implementaciones (Hadoop, Storm, Spark, etc.), así como metodologías de almacenamiento en la nube y computación. Este conocimiento los preparará para manejar grandes conjuntos de datos y realizar análisis estadísticos y computacionales complejos de estos datos, para obtener resultados en la identificación, predicción, previsión, simulación u optimización

AUDIENCIA

Personal de Análisis en Banca, Seguros, estudiantes universitarios, público en general. Licenciados en Física, Estadística, Matemáticas, Ciencias Económicas y Financieras, e Ingenierías.

PRE REQUISITOS

- Preferiblemente con base y conocimientos en: Bases de datos, estadística, Linux, R, programación, algoritmos, SQL.

OBJETIVOS

- Adquisición de conocimientos y comprensión de base suficientes para desarrollar y aplicar ideas originales en contextos

diversos.

- Aplicación de sus conocimientos y su comprensión, así como sus habilidades para resolver problemas, en entornos nuevos o no familiares y en contextos amplios, empleando para ello información y recursos publicados por profesionales de otros campos y accesibles mediante plataformas de contenidos digitales, archivos, bibliotecas o (si fuese posible) mediante el diálogo directo con dichos profesionales.
- Comunicar y fundamentar sus conclusiones, conocimientos y el marco conceptual en que se basan, de manera clara y sin ambigüedades, tanto a interlocutores expertos como no expertos, utilizando preferentemente para ello tecnologías digitales.
- Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
- Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
- Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios
- Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades
- Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CERTIFICACIÓN DISPONIBLE

Certificación emitida por COGNOS.

CONTENIDO

1. MODULO I. CONCEPTOS Y METODOS PARA ANALISIS DE BIG DATA

- 1.1. DATOS ESTRUCTURADOS, SEMI-ESTRUCTURADOS Y NO ESTRUCTURADOS
- 1.2. COMPUTACION PARALELA Y DISTRIBUIDA. MAPREDUCE
- 1.3. TECNOLOGIAS CLOUD COMPUTING Y ALMACENAMIENTO DE ALTO RENDIMIENTO
- 1.4. METODOS Y ARQUITECTURAS DE PROCESAMIENTO DE DATOS

2. MODULO II. PROCESAMIENTO PARALELO Y DISTRIBUIDO DE DATOS

- 2.1. ANALISIS DE BIG DATA EN R Y PYTHON
- 2.2. ECOSISTEMA APACHE HADOOP

3. MODULO III. BASES DE DATOS NO RELACIONALES (NOSQL)

- 3.1. METODOLOGIAS NOSQL

4. MODULO IV. PROCESAMIENTO DE DATOS EN MEMORIA Y OTRAS TECNICAS AVANZADAS

- 4.1. PROCESAMIENTO DE DATOS EN MEMORIA
- 4.2. TECNOLOGIAS AVANZADAS DE PROCESAMIENTO DISTRIBUIDO DE DATOS

5. MODULO V. CASO PRACTICO

5.1. ANALISIS DE UN CASO DE LA PLATAFORMA WWW.KAGGLE.COM

★ BENEFICIOS

Al finalizar el curso los participantes tendrán los conocimientos y comprensión de base suficientes para desarrollar y aplicar ideas originales en contextos diversos.